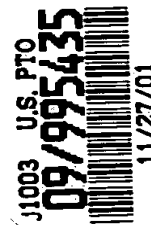




Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets



Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

00403386.6

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN  
THE HAGUE, 22/08/01  
LA HAYE, LE

**THIS PAGE BLANK (USPTO**



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

**Blatt 2 der Bescheinigung**  
**Sheet 2 of the certificate**  
**Page 2 de l'attestation**

Anmeldung Nr.:  
Application no.:  
Demande n°: 00403386.6

Anmeldetag:  
Date of filing:  
Date de dépôt: 01/12/00

Anmelder:  
Applicant(s):  
Demandeur(s):  
Koninklijke Philips Electronics N.V.  
5621 BA Eindhoven  
NETHERLANDS

Bezeichnung der Erfindung:  
Title of the invention:  
Titre de l'invention:  
Method and device for video scene composition from varied data

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:  
State:  
Pays:

Tag:  
Date:  
Date:

Aktenzeichen:  
File no.  
Numéro de dépôt:

Internationale Patentklassifikation:  
International Patent classification:  
Classification internationale des brevets:  
/

Am Anmeldetag benannte Vertragsstaaten:  
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR  
Etats contractants désignés lors du dépôt:

Bemerkungen:  
Remarks:  
Remarques:

**THIS PAGE BLANK** (USPTO

**FIELD OF THE INVENTION**

The present invention relates to a method of composing an MPEG-4 video scene content at least from a first set of input video objects coded according to the MPEG-4 standard, said method comprising a first decoding step for generating a first set of decoded MPEG-4 video objects from said first set of input video objects, and a rendering step for generating composed frames of said video scene from at least said first set of decoded MPEG-4 video objects.

For example, this invention may be used in the field of digital television broadcasting and implemented in a set top box as an Electronic Program Guide (EPG).

**BACKGROUND OF THE INVENTION**

The MPEG-4 standard relative to system aspects, referred to as ISO/IEC 14496-1, provides functionality for multimedia data manipulation. It is dedicated to scene composition containing different natural or synthetic objects, such as two or three dimensional images, video clips, audio tracks, texts or graphics. This standard allows scene content creation usable with multiple applications, allows flexibility in object combination, and offer means for user-interaction in scenes containing multiple objects. This standard may be used in a communication system comprising a server and a client terminal via a communication link. In such applications, MPEG-4 data exchanged between both sets are streamed on said communication link and exploited on the client terminal to create multimedia applications.

The international patent application WO 00/01154 describes a terminal and method of the above kind for composing and presenting MPEG-4 video programs. This terminal comprises :

- a terminal manager for managing the overall processing tasks,
- decoders for providing decoded objects,
- a composition engine for maintaining, updating and assembling a scene graph of the decoded objects,
- a presentation engine for providing a scene for presentation.

**SUMMARY OF THE INVENTION**

It is an object of the invention to provide a cost-effective and optimised method of video scene composition that allows the composition of an MPEG-4 video scene simultaneously from video data coded according to the MPEG-4 video standard referred to as ISO/IEC 14496-2 and video data coded according to other video standards. The invention takes the following aspects into consideration.

The composition method according to the prior art allows the composition of a video scene from a set of decoded video objects coded according to the MPEG-4 standard. To this end, a composition engine maintains and updates a scene graph of the current objects, including their relative position in a scene and their characteristics, and provide a corresponding  
5 list of objects to be displayed to a presentation engine. In response, the presentation engine retrieves the corresponding decoded object data that is stored in respective composition buffers. The presentation engine renders the decoded objects for providing a scene for presentation on a display.

With the large use of digital networks such as Internet, most of multimedia  
10 applications resulting in a video scene composition collects video data from different sources to enrich their content. In this context, if this prior art method is used for a video scene composition, collected data not compliant with the MPEG-4 standard could not be rendered, which could lead to a poor video scene content or produce an error in the applications. Indeed, this prior art method is very restrictive since the video scene composition can exclusively be  
15 performed from video objects coded according to the MPEG-4 system standard, which excludes the use of other video data in the video scene composition, such as MPEG-2 video data.

To solve the limitations of the prior art method, the method of video scene composition according to the invention is characterized in that it comprises :  
20 a) a second decoding step for generating a set of decoded video data from a second set of input video data not MPEG-4 compliant,  
b) a video object creation step for generating a second set of video objects, each created video object being formed by the association of a decoded video data extracted from said set of decoded video data, and a set of properties for defining characteristics of said  
25 decoded video data in the video scene, said second set of video objects being rendered with said first set of decoded MPEG-4 video objects during said rendering step.

This allows the rendering of the overall input video objects in the scene for resulting in an MPEG-4 video scene.  
30

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereinafter.

## 35 **BRIEF DESCRIPTION OF THE DRAWINGS**

The particular aspects of the invention will now be explained with reference to the embodiments described hereinafter and considered in connection with the accompanying drawings, in which identical parts or sub-steps are designated in the same manner :

Fig.1 depicts the different functional blocks of the MPEG-4 video scene composition according to the invention,

Fig.2 depicts the hardware implementation of the MPEG-4 video scene composition method according to the invention,

5 Fig.3 depicts an embodiment of the invention.

### **DETAILED DESCRIPTION OF THE INVENTION**

The invention allows a video scene composition from input video streams encoded according to the MPEG-4 standard and input video streams coded according to other video standards different than the MPEG-4 standard. It is described in the case where said video streams coded according to other video standards different than the MPEG-4 standard correspond to video streams coded according to the MPEG-2 video standard, but it would be apparent to a person skilled in the art that this invention can also be used with other standards such as H.263, MPEG-1 or proprietary company format.

Figure 1 depicts the different functional blocks of the video scene composition according to the invention.

The method of scene composition according to the invention comprises the following functional steps:

- 1) a first decoding step 101 for decoding an input video stream 102 containing input video objects coded according to the MPEG-4 video standard. This decoding step 101 results in decoded MPEG-4 video objects 103. If the input video stream 102 corresponds to a demultiplexed video stream or comprises a plurality of elementary video streams, each elementary video stream is decoded by a separate decoder during the decoding step 101.
- 2) a second decoding step 104 for decoding an input video stream 105 containing input coded video data not coded according to the MPEG-4 video standard, but coded for example according to the MPEG-2 video standard. This decoding step results in decoded MPEG-2 video data 106. If the input video stream 105 corresponds to a demultiplexed video stream or comprises a plurality of elementary video streams, each elementary video stream is decoded by a separate decoder during the decoding step 104.
- 3) a video object creation step 107 for generating video objects 108 from said decoded MPEG-2 video data 106. This step consists in associating to each decoded video data 106, a set of properties defining its characteristics in the final video scene. Each data structure, linked to a given video data 106, comprises for example:
  - a) a field "depth" for defining the depth of said video data in the video scene (e.g. first ground or second ground),
  - b) a field "transform" for defining a geometric transform of said video data (e.g. a rotation characterised by an angle),

c) a field "transparency" for defining the transparency coefficient between said video data and other video objects in the video scene.

By this way, the resulting video objects 108 are compatible with MPEG-4 video objects 103 in the sense that each video object 108 not only contain video frames but refer to a set of characteristics allowing its description in the video scene.

4) a rendering step 109 for assembling video objects 103 and 108. To this end, video objects 103 and 108 are rendered in using their own object properties, or in using object properties (filled during the video object creation step 107, for video objects 103), contained in a BIFS stream 111 (Binary Format for Scene), said BIFS stream 111 containing scene graph description describing each object properties in the scene. The assembling order of video objects is determined by the depth of each video object to be rendered : the video objects composing backgrounds are first assembled, then the video objects composing foregrounds are finally assembled. This rendering results in the delivering of an MPEG-4 video scene 110.

As an example, in an electronic program guide (EPG) allowing a viewer to browse TV programs, this method can be used for composing a video scene from an MPEG-2 video stream 105 and an MPEG-4 video stream 102, said MPEG-2 video stream 105 defining after decoding 104 a full screen background MPEG-2 video, while said MPEG-4 video stream defining after decoding 101 a first object MPEG4\_video\_object1 corresponding to a video of reduced format (used as TV preview for example) and a second object MPEG4\_video\_object2 corresponding to textual information (used as time and channels indications).

The rendering of these three video elements is made possible by the association of a set of properties Scene\_video\_object3 to the decoded MPEG-2 video in order to define the characteristics of this MPEG-2 video in the video scene, this association resulting in the video object MPEG4\_video\_object3. Concerning the two decoded MPEG-4 objects, they are associated each one, according to the MPEG-4 syntax relative to scene description, to a set of properties Scene\_video\_object1 (and Scene\_video\_object2) in order to define their characteristics in the video scene. These two sets Scene\_video\_object1 and Scene\_video\_object2 may be filled by pre-set parameters or by parameters contained in the BIFS stream 111. In this last possibility, the composed scene can be real-time updated, especially if the BIFS update mechanism, well known by people skilled in the art, is used, which allows to change the characteristics of video objects in the scene.

In each video object structure, a structure Buffer\_video is also defined to access video data, i.e. video frames, by three pointers pointing on component Y, U and V of each video data. For example, the component Y of the video object 1 is accessed by pointer pt\_video1\_Y, while the components U and V are accessed by pointers pt\_video1\_U and pt\_video1\_V respectively.



5

The corresponding scene graph has the following structure :

```

Scene_graph {
    MPEG4_video_object1 {
        Scene_video_object1 {
            depth1
            transform1
            transparency1
        }
        Buffer_video1 {
            pt_video1_Y
            pt_video1_U
            pt_video1_V
        }
    }
    MPEG4_video_object2 {
        Scene_video_object2 {
            depth2
            transform2
            transparency2
        }
        Buffer_video2 {
            pt_video2_Y
            pt_video2_U
            pt_video2_V
        }
    }
    MPEG2_video_object3 {
        Scene_video_object3 {
            depth3
            transform3
            transparency3
        }
        Buffer_video3 {
            pt_video3_Y
            pt_video3_U
            pt_video3_V
        }
    }
}

```

The rendering step 109 first assembles the MPEG-4 objects MPEG4\_video\_object1 and MPEG4\_video\_object2 in a composition buffer in respecting characteristics of structures

Scene\_video\_object1 and Scene\_video\_object2. Then, the video object MPEG2\_video\_object3 is rendered with previously rendered MPEG-4 objects in taking into account the characteristics of the structure Scene\_video\_object3.

5

Figure 2 depicts the hardware architecture 200 for implementing the different steps of the video scene composition according to the invention.

This architecture is structured around a data bus 201 to ensure data exchange between the different processing hardware units. This architecture includes an input peripheral 202 for receiving MPEG-4 and MPEG-2 input video streams that are both stored in the mass storage 203.

The decoding of video streams coded according to the MPEG-4 standard is done with the signal processor 204 (referred to as SP in the figure) in executing instructions relative to an MPEG-4 decoding algorithm stored in memory 205, while the decoding of video streams coded according to MPEG-2 is also done with the signal processor 204 in executing instructions relative to an MPEG-2 decoding algorithm stored in said memory 205 (or an appropriate decoding algorithm if the input video stream is coded according to a video standard different than the MPEG-2 one). Once decoded, MPEG-4 video objects are stored in a first data pool buffer 206, while MPEG-2 video data are stored in a second data pool buffer 211.

The video rendering step is performed by the signal processor 204 in executing instructions relative to a rendering algorithm stored in the memory 205. The rendering is performed in assembling in a composition buffer 210 not only decoded MPEG-4 objects but also decoded MPEG-2 data. To this end, in order to avoid multiple and expensive data manipulation, decoded MPEG-2 data are recopied by a signal co-processor 209 (referred to as SCP in the figure) directly from buffer 211 to said composition buffer 210. This recopy ensures that a minimum computational load is used, which does not limit other tasks in the application such as the decoding or the rendering tasks. In the same time, the set of properties relative to said MPEG-2 data is filled and taken into account by the signal processor during the rendering step. By this way, MPEG-2 data have a similar structure than MPEG-4 ones (i.e. association of video data and properties), which allows the rendering of the overall input video objects. Thus, the rendering takes into account not only MPEG-4 objects properties and MPEG-2 properties, but also data relative :

- 1) to the action of a mouse 207 and/or a keyboard 208,
  - 2) and/or to BIFS commands issued from a BIFS Stream stored in the storage device 203 or received via input peripheral 202,
- for changing the position of video objects in the video scene being built, according to the action of the viewer using the EPG.

When a rendered frame is available in the content of buffer 210, it is presented to an output video peripheral 212 for being displayed on a display 213.

In this implementation, the processor 204 and the co-processor 209 are simultaneously used so that MPEG-4 input video objects composing the next output frame of the video scene can ever been decoded during the recopy by the SCP in the composition buffer, of decoded MPEG-2 video data composing the current output frame of the video scene. This is made possible by the non CPU-consuming process (Clock Pulse Units) done by the SCP which allows the SP to use all the CPU processing capacity. This optimised processing will be highly appreciated by people skilled in the art, especially in a real-time video scene composition context where input video objects of large size, requiring high computational resources, have to be processed.

Figure 3 depicts an embodiment according to the invention. This embodiment corresponds to an electronic program guide application (EPG) allowing a viewer to watch on a display 304 varied information relative to TV channels programs. To this end, the viewer navigates through the screen in translating, by means of a mouse-like/pointer 305 device, the browsing window 308 in a channels space 306 and time space 307, said browsing window playing the corresponding video preview of the chosen time/channel combination. The browsing window 308 is overlaid and blended on top of a background video 309.

The different steps according to the invention described in figures 1 are implemented in a set-top box unit 301 receiving input video data from an outside world 302. Said input video data corresponds for example in this example to MPEG-4 video data delivered by a first broadcaster (e.g. video objects 306-307-308), and to MPEG-2 video data delivered by a second broadcaster (e.g. video data 309), via a communication link 303. Said input video data are processed according to the different steps of the invention described in figure 1 in using an hardware architecture as described in figure 2, resulting in MPEG-4 video composed frames composed by the overall input video objects.

Of course, the presented graphic designs does not restrict the scope of the invention since other graphic designs can be envisaged without deviating from the scope of the invention.

There has been described an improved method of composing a scene content simultaneously from input video streams encoded according to the MPEG-4 video standard and according to non MPEG-4 compliant video data (i.e. not coded according to the MPEG-4 standard) such as MPEG-2 video data. The method according to the invention relies on a video object creation step allowing to compose an MPEG-4 video scene from said non MPEG-4 compliant video data, thanks to the association of scene properties to said non MPEG-4 compliant video data.

Of course, this invention is not restricted to the presented structure of scene properties associated to said non MPEG-4 video data, and other fields defining this structure can be considered without deviating from the scope of the invention.

5           This invention can be implemented in several manners, such as by means of wired electronic circuits or alternatively, by means of a set of instructions stored in a computer-readable medium, said instructions replacing at least a part of said circuits and being executable under the control of a computer, a digital signal processor or a digital signal co-processor in order to carry out the same functions as fulfilled in said replaced circuits. The invention then  
10 also relates to a computer-readable medium comprising a software module that includes computer executable instructions for performing the steps, or some steps, of the above described method.

**CLAIMS**

1. A method of composing an MPEG-4 video scene content at least from a first set of input video objects coded according to the MPEG-4 standard, said method comprising a first decoding step for generating a first set of decoded MPEG-4 video objects from said first set of input video objects, and a rendering step for generating composed frames of said video scene from at least said first set of decoded MPEG-4 video objects, **characterized in that** said method also comprises :
- 5
- a) a second decoding step for generating a set of decoded video data from a second set of input video data not MPEG-4 compliant,
- 10 b) a video object creation step for generating a second set of video objects, each created video object being formed by the association of a decoded video data extracted from said set of decoded video data, and a set of properties for defining characteristics of said decoded video data in the video scene, said second set of video objects being rendered with said first set of decoded MPEG-4 video objects during said rendering step.
- 15
2. A method of composing an MPEG-4 video scene content as claimed in claim 1, **characterized in that** said properties define the depth, a geometric transform and the transparency coefficient.
- 20
3. A method of composing an MPEG-4 video scene content as claimed in claim 1, **characterized in that** said second decoding step is dedicated to the decoding of input video data coded according to the MPEG-2 video standard.
- 25
4. A set-top box product for composing an MPEG-4 video scene at least from a first set of input video objects coded according to the MPEG-4 standard, said set-top box comprising a first decoding means for generating a first set of decoded MPEG-4 video objects from said first set of input video objects, and rendering means for generating in a composition buffer composed frames of said video scene from at least said first set of decoded MPEG-4 video objects, **characterized in that** said method also comprises :
- 30 a) a second decoding means for generating a set of decoded video data from a second set of input video data not MPEG-4 compliant,
- b) video object creation means for generating a second set of video objects, each created video object being formed by the association of a decoded video data extracted from said set of decoded video data, and a set of properties for defining characteristics of said
- 35 decoded video data in the video scene, said second set of video objects being rendered with said first set of decoded MPEG-4 video objects by said rendering means.

5. A set-top box product as claimed in claim 4, **characterised in that** :
- a) decoding means correspond to the execution of dedicated program instructions by a signal processor, said program instructions being loaded in said signal processor or in a memory,
  - b) video object creation means correspond to the execution of dedicated program instructions by said signal processor, said program instructions being loaded in said signal processor or in a memory, said signal processor being dedicated to the association of data defining properties, to each video data constituting said set of decoded video data, for defining characteristics of each decoded video data in the video scene,
  - c) rendering means not only correspond to the execution of dedicated program instructions by said signal processor, said program instructions being loaded in said signal processor or in a memory, but also to the execution of hardware functions by a signal co-processor in charge of the recopy of said second set of video objects in said composition buffer.
6. A set-top box product as claimed in claim 4, **characterised in that** it comprises means for taking into account user-interactions, for modifying the relative spatial position of said first set of decoded MPEG-4 video objects and said second set of video objects in the MPEG-4 video scene.
7. A set-top box product as claimed in claim 4, **characterized in that** said second decoding means are dedicated to the decoding of input video data coded according to the MPEG-2 video standard.
8. Computer program product, for a device composing an MPEG-4 video scene from MPEG-4 video objects and non MPEG-4 video objects, that comprises a set of instructions, which, when loaded into said device, causes said device to carry out the method as claimed in claims 1 to 3.

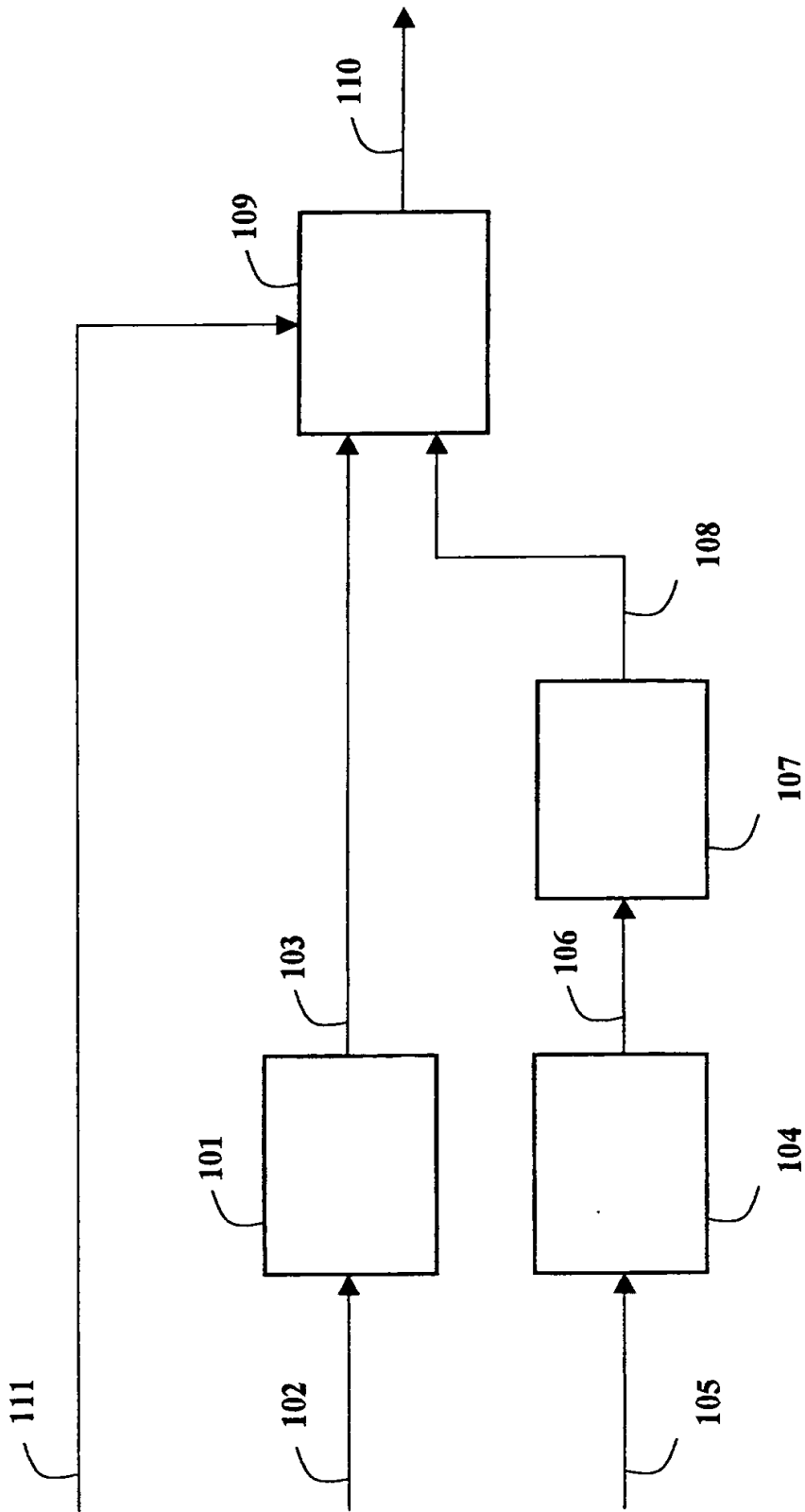


FIG.1

2 / 3

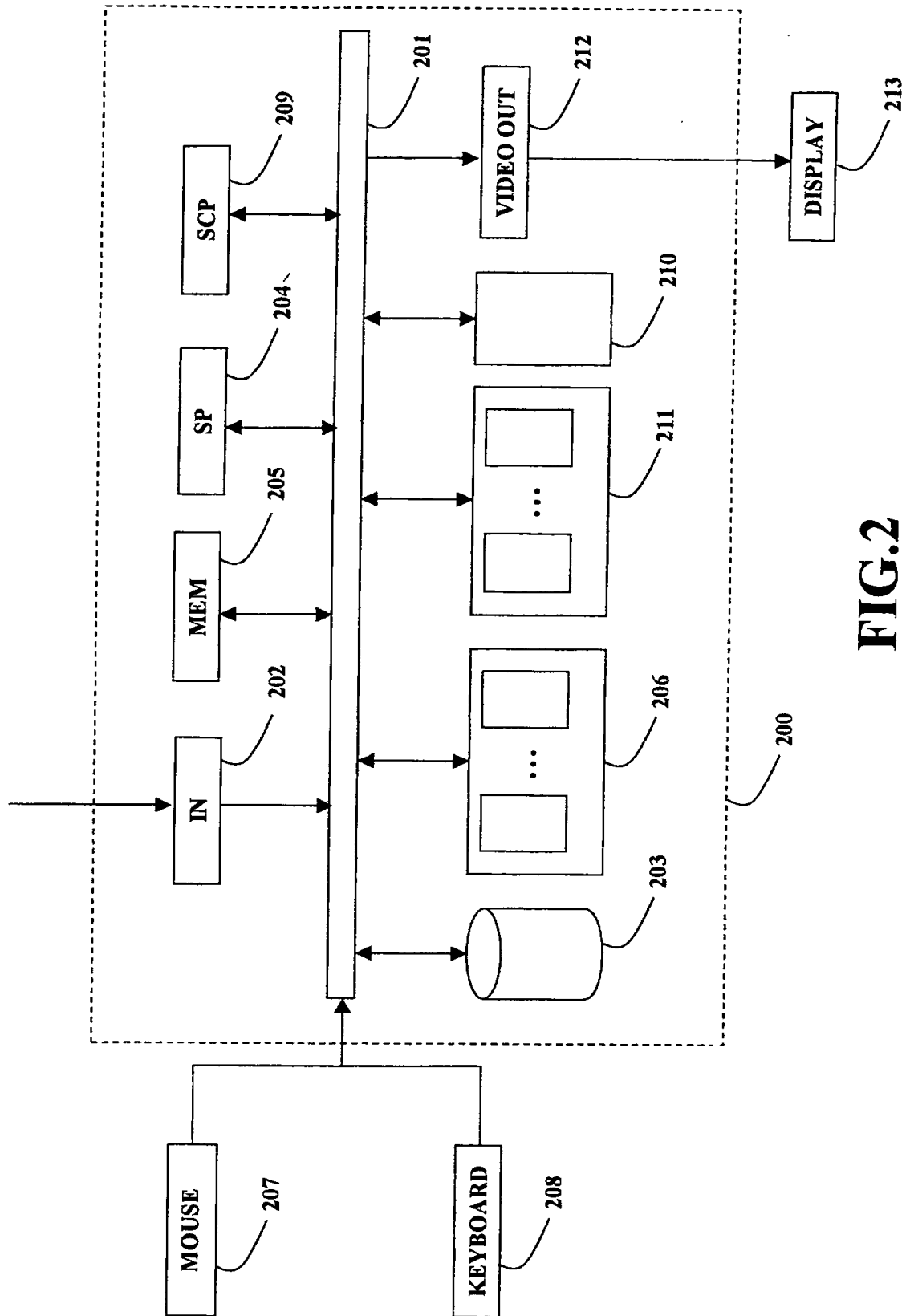
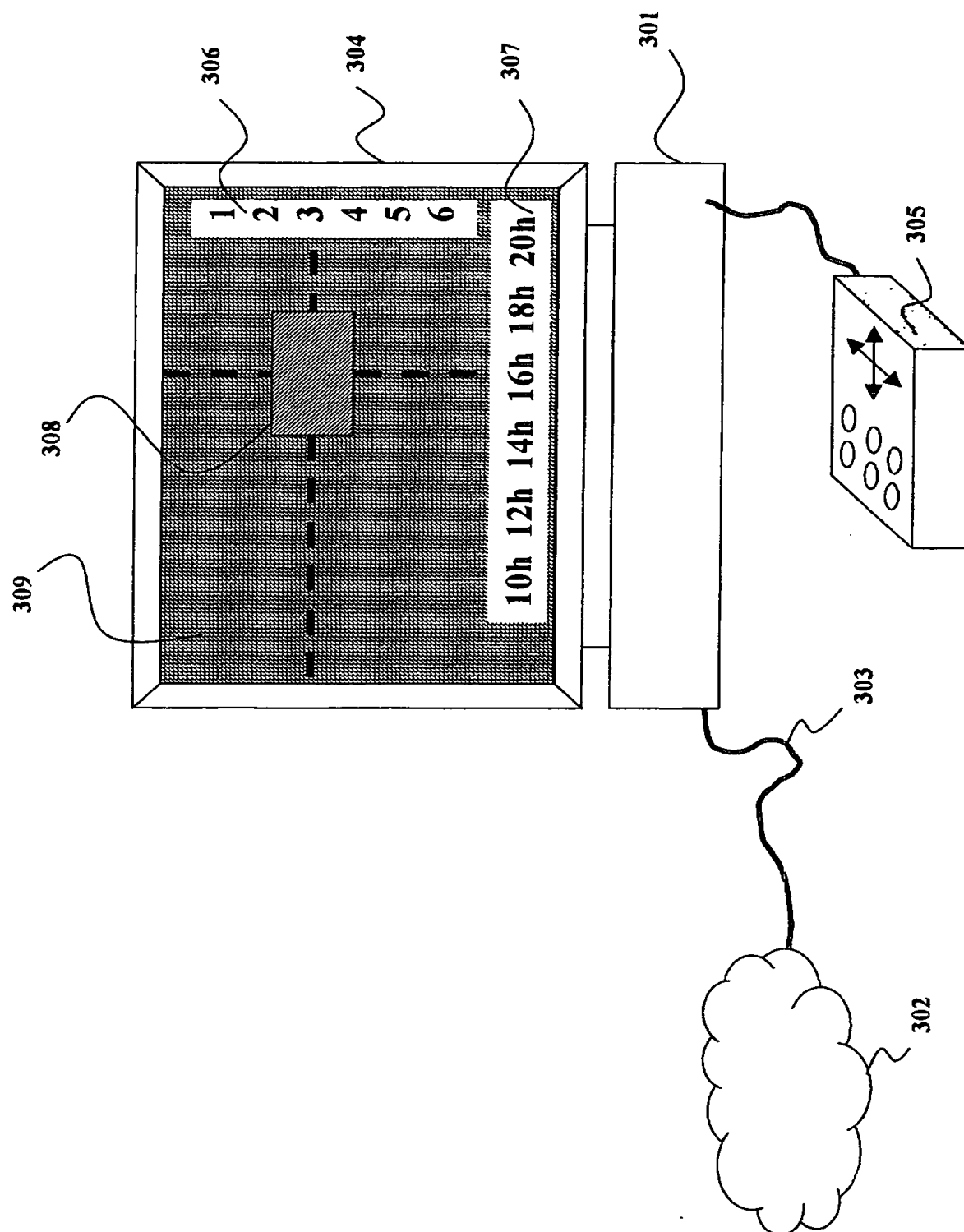


FIG.2



3/3



**THIS PAGE BLANK (USPTO)**